the reference (original) version against which both the "A" and "B" versions are to be compared and graded. One of "A" and "B" is a processed (i.e. coded and decoded) version and the other is a hidden reference, identical to the "Ref" version.

You are not told which of "A" or "B" is the processed version and which is the hidden reference and this will change randomly from one trial to the next. You will hear each triplet twice (i.e. "Ref" "A" "B" -- "Ref" "A" "B") prior to scoring "A" and "B". This should allow a detailed comparison between "Ref", "A" and "B".

You are asked to judge the "Basic Audio Quality" of the "A" and "B" versions in each trial. This attribute, quality, is related to any and all differences between the reference and the coded/decoded program excerpt. NOTE: Any difference between the reference and the coded/decoded program is to be considered an impairment.

It is not possible to list all possible differences that may be created by the form of sound signal processing being evaluated in these tests. However what follows is a list of the main differences that may be expected.

It includes such things as harmonic distortions, added 'pops' or 'cracks', quantization noise in subbands, pre-echoes (or other time smearing effects), changes in loudness, changes in timbre, changes in spatial presentation, changes in background noise or reverberance. Anything else that the listener detects as a difference must be included in their overall rating.

In each trial you are asked to rate the perceived difference (if any) between "Ref" and "A" and also the difference between "Ref" and "B" using the grading scale:

~ 5.0 Imperceptible

~ 4.0 Perceptible, but not annoying

~ 3.0 Slightly Annoying

~ 2.0 Annoying

~ 1.0 Very Annoying

The grading scale is to be considered as a continuous equal-interval scale with descriptions at these five anchor points" which define specific values.

Two grades must be given on each trial, one for "A" and one for "B". <u>At least one grade of "5" must be given on each trial</u> since one of "A" or "B" is the hidden reference, and therefore is imperceptible by definition.

Please mark your grades at the end of the trial, preferably to just one decimal place. It should be noted that the order of presentation of the test blocks and the position of the hidden reference is randomized and reordered for each test session.

3.0 Confidentiality

You may not discuss what you did here until after the report goes to Committee and is made public. If you would like a copy of the report be sure to sign up on the document mailing list.

ANNEX E

Statistical Procedures

This statistical analysis was performed using Analysis of Variance of the difference scores between the reference and the coder. The results were confirmed by categorical analysis, using 0 for scores where the coder was rated above the reference and 1 for scores where the coder was rated the same or below the reference. The resulting Chi-squared tests were not as sensitive to differences as ANOVA. The residuals were evaluated for symmetry and normality. The residual tests always failed normality, usually for too heavy tails. The distribution was symmetrical, however. Wilcoxon/Kruskal-Wallis tests, non-parametric tests similar to the one-way Analysis of Variance, which only assumes symmetry, were performed with similar results. The variance, as a rule, was non-homogenous across selections. On a selection by selection basis, the average of the difference was calculated and the average tested for a difference from zero (no artifacts in the coder) by both a t-test and a signed rank test. The resulting averages, standard deviations, 95% confidence intervals and probability levels (a difference is real if the number is small, usually less than .05 or .01) have been presented. The probability test is a one-way test (Is the average less than zero?)

Since quite a few scores for the reference were lower than those for the coder, the data were also analyzed using only listeners who correctly identified the reference (or who scored both the reference and coder the same) at least 75% of the time (75% is the threshold used in signal detection and just- noticeable difference tasks). The number of listeners meeting these criteria across the four tests ranged from only two (2) to a maximum of seven (7) out of 22 listeners.

Summary of Additional Analysis

It should be noted that when making comparisons at the 5% error/95% confidence level, false positives and spurious results due to random chance will occur. It is best to test each individual segment of the test segment at 1% (overall confidence level with ten test selections of 90%) or at 0.5% (overall confidence level with ten test selections of 95%).

Another point worth noting is that since the coder is never allowed to be judged better than the reference, the correct error probability is twice that of a two-sided test and is one-sided. This means that all the 95% confidence intervals are really 97.5% confidence intervals when testing whether a segment has been correctly detected (i.e. different from 0, or, not different from the reference).

In addition to the above test results, such things as age group, listener position in the room, A/B voting preference and fatigue versus learning curve were examined.

It was found that seating position in the room was not a differentiating factor (although many of the listeners would swear that there were favored positions).

It was found that removing non-reliable judgments (when a listener judged the same test selection differently by 2 or more grades), made just about no difference at all in the final scores.

Somewhat related is the learning curve or fatigue factor. In examining the data for a systematic change in either the mean or the standard deviation for the diffference scores and the percent correct detection, there are very few statistically significant shifts found (only on

Berlioz which showed a shift in the mean but not standard deviation the new coder, and showed a shift in percent correct on the old one).

Interestingly there was no evidence of reduced variability for the 1993 coder, but the new coder turned up lower standard deviations on quite a few of the selections (9 of 10 in the mixdown and multichannel tests [probability = .012] and 8 of 10 in the stereo test [probability = .067]).

In checking for a preference to vote "A", the percent of correct responses as a function of whether the coder appeared in the A or B slot showed a real preference for A as expected except in the stereo test.

Age groups of < 35 years, 35 - 44, 45 - 46 and > 47 showed no differences except in stereo in which the 2 oldest groups did the best (>45 years had 73% correct).

Comparing accuracy of the first trial versus second trial (same selection) no correlation at all was found. This is unexpected since one would expect some learning or some fatigue to have been a factor.

Part VI

Matrix of Tests
&
Index of Tasks for Individual Tests

for

digital HDTV
Grand Alliance System

Conducted by/with

Advanced Television Test Center
Advanced Television Evaluation Laboratories
Cable Television Laboratories
Task Force on Digital-Specific Tests (SS/WP-2)
Task Force on Audio (SS/WP-2)
Hitachi America Ltd.
IBM

(April - October, 1995)

Matrix of Tests on the *digital* HDTV Grand Alliance System

see page VI-1-13 for Legend of terms)

All tests were conducted in accordance with the Grand Alliance System Test Procedures; SSWP2-1306, except as noted.

Transmission Tests

Power Measurement

ATTC Test #	TEST	SUB-TEST			NOTES	TEST PLAN
	Power Measurement	OBJ	REC	EO&C		SECTION
229	Peak/Average Power Meas't Generic Method & Boonton Instrument	✓ ✓				I-3.2

Co-Channel Transmission Tests

ATTC Test #	TEST	D	SUB-TEST							NOTES	TEST PLAN
		LVL	TOV	CCIR 3	POU	POR	POF	RNG	EO&C		SECTION
57	RN-I/N	-25 dBm	✓							†	I-3.7.3.1
16 TOV 235 CCIR3	CO-A/N	M	✓	✓					✓		I-3.7.3.1
16 TOV 235 CCIR3	CO-A/N	W	✓	✓	✓			✓	✓		I-3.7.3.1
17	CO-N/A	M	✓*							BER method	I-3.7.3.1
17	CO-N/A	W	✓		✓	✓	✓			BER method plus viewing	I-3.7.3.1
56	CO-N/A Δ Frequency Offset	W	✓*							BER method	I-3.7.3.1
18	CO-A/A	M	✓*							BER method	I-3.7.3.1
237	CO-A/A Δ Delay	M	✓*							BER method	I-3.7.3.1
18	CO-A/A	W	✓*							BER method	I-3.7.3.1
237	CO-A/A Δ Delay	W	✓*							BER method	I-3.7.3.1
264	CO-A/A Δ Frequency Offset	W	✓*							BER method	I-3.7.3.1
265	CO-A/A Δ Frequency Offset (Δ Delay)	W	✓*							BER method	I-3.7.3.1
235	BTSC Audio for CO-A/N	W							✓	EO&C for 6 TV Sets @ CCIR3 Video Level	I-4
All above tests use 1080I x 1920 pixel Scan Format											

† The RN-I/N calibration test was performed once per day and once for each different panel of expert observers, as a training exercise.

✓ Data taken in test plan for Grand Alliance system.

* When a TOV was determined for interference/impairment into ATV, acquisition at TOV was confirmed.

Transmission Tests

ATTC Test #	TEST	D	SUB-TEST							NOTES	TEST PLAN
		LVL	TOV	CCIR 3	POU	POR	POF	RNG	EO &C		SECTION
2 TOV 3 CCIR3	UP-A/N	-25 dBm	✓	✓‡	✓			✓	✓		I-3.7.3.2
2 TOV 3 CCIR3	UP-A/N	M	✓	✓‡					✓		I-3.7.3.2
2 TOV 3 CCIR3	UP-A/N	W	✓	✓‡	✓			✓	✓		I-3.7.3.2
4	UP-N/A	S	✓*							BER method plus viewing	I-3.7.3.2
4	UP-N/A	M	✓*							BER method	I-3.7.3.2
4	UP-N/A	W	✓*							BER method plus viewing	I-3.7.3.2
6	UP-A/A	S	✓*							BER method	I-3.7.3.2
6	UP-A/A	M	✓*							BER method	I-3.7.3.2
6	UP-A/A	W	✓*							BER method	I-3.7.3.2

ATTC Test #	TEST	D	SUB-TEST							NOTES	TEST PLAN
		LVL	TOA	Stereo CCIR 4	Stereo CCIR 3	SAP CCIR 4			EO &C		SECTION
3	BTSC Audio for UP-A/N Stereo/ Mono, & SAP	S		✓	✓	✓			✓	EO&C for 24 TV Sets Stereo/Mono and 15 TV sets SAP	I-4
		M		✓	✓	✓			✓		
		W			✓				✓		
		W 7dB V/A Ratio			✓				✓	3 receivers for 7dB Visual/Aural ratio *This group of tests essentially supplementary to original test plan*	
All above tests use 1080I x 1920 pixel Scan Format											

✓ Data taken in test plan for Grand Alliance system.

* When only TOV was determined for interference/impairment into ATV, acquisition at TOV was confirmed.

‡ = CCIR3 interference levels determined for all 24 receivers. *Extension from 12+ to 24 receivers supplementary to original test plan*

Transmission Tests

ATTC Test #	TEST	D LVL	SUB-TEST TOV	CCIR 3	POU	POR	POF	RNG	EO &C	NOTES	TEST PLAN SECTION
9 TOV 10 CCIR3	LO-A/N	-25 dBm	✓	✓	✓			✓	✓		I-3.7.3.2
9 TOV 10 CCIR3	LO-A/N	M	✓	✓					✓		I-3.7.3.2
9 TOV 10 CCIR3	LO-A/N	W	✓	✓‡	✓			✓	✓		I-3.7.3.2
11	LO-N/A	S	✓*							BER method plus viewing	I-3.7.3.2
11	LO-N/A	M	✓*							BER method	I-3.7.3.2
11	LO-N/A	W	✓*							BER method plus viewing	I-3.7.3.2
13	LO-A/A	S	✓*							BER method	I-3.7.3.2
13	LO-A/A	M	✓*							BER method	I-3.7.3.2
13	LO-A/A	W	✓*							BER method	I-3.7.3.2

ATTC Test #	TEST	D LVL	SUB-TEST TOA		Stereo CCIR 3	SAP CCIR 3			EO &C	NOTES	TEST PLAN SECTION
10	BTSC Audio for LO-A/N Stereo/ Mono, & SAP.	W			✓	✓			✓	EO&C for 6 TV Sets Stereo/Mono and SAP *This group of tests supplementary to original test plan*	
All above tests use 1080I x 1920 pixel Scan Format											

✓ Data taken in test plan for Grand Alliance system.

* When only TOV is determined for interference/impairment into ATV, acquisition at TOV was confirmed.

‡ = CCIR3 interference levels determined for all 24 receivers. *Extension from 12+ to 24 receivers supplementary to original test plan*

Transmission Tests

ATTC Test #	TEST	D LVL	SUB-TEST								NOTES	TEST PLAN SECTION
			TOV	CCIR 4	CCIR 3	POU	POR	POF	RNG	EO &C		
263	RN-I/N	S	✓								†	I-3.7.3.3
20	N-8 Taboo A/N	S	✓	✓						✓		I-3.7.3.3
20	N-8 Taboo A/N	M	✓	✓						✓		I-3.7.3.3
20	N-8 Taboo A/N	W	✓		✓					✓		I-3.7.3.3
248	N-3 Taboo A/N	S	✓	✓						✓		I-3.7.3.3
248	N-3 Taboo A/N	M	✓	✓						✓		I-3.7.3.3
248	N-3 Taboo A/N	W	✓		✓					✓		I-3.7.3.3
28	N-2 Taboo A/N	S	✓	✓						✓		I-3.7.3.3
28	N-2 Taboo A/N	M	✓	✓						✓		I-3.7.3.3
28	N-2 Taboo A/N	W	✓		✓					✓		I-3.7.3.3
32	N+2 Taboo A/N	S	✓	✓						✓		I-3.7.3.3
32	N+2 Taboo A/N	M	✓	✓						✓		I-3.7.3.3
32	N+2 Taboo A/N	W	✓		✓					✓		I-3.7.3.3
249	N+3 Taboo A/N	S	✓	✓						✓		I-3.7.3.3
249	N+3 Taboo A/N	M	✓	✓						✓		I-3.7.3.3
249	N+3 Taboo A/N	W	✓		✓					✓		I-3.7.3.3
36	N+4 Taboo A/N	S	✓	✓						✓		I-3.7.3.3
36	N+4 Taboo A/N	M	✓	✓						✓		I-3.7.3.3
36	N+4 Taboo A/N	W	✓		✓					✓		I-3.7.3.3
All above tests use 1080I x 1920 pixel Scan Format												

† The RNI/N calibration test was performed once per day, and once for each different panel of expert observers, as a training exercise.

✓ Data taken in test plan for Grand Alliance system.

Transmission Tests

ATTC Test #	TEST	D LVL	SUB-TEST								NOTES	TEST PLAN SECTION
			TOV	CCIR 4	CCIR 3	POU	POR	POF	RNG	EO &C		
44	N+8 Taboo A/N	S	✓	✓						✓		I-3.7.3.3
44	N+8 Taboo A/N	M	✓	✓						✓		I-3.7.3.3
44	N+8 Taboo A/N	W	✓		✓					✓		I-3.7.3.3
48	N+14 Taboo A/N	S	✓	✓						✓		I-3.7.3.3
48	N+14 Taboo A/N	M	✓	✓						✓		I-3.7.3.3
48	N+14 Taboo A/N	W	✓		✓‡					✓		I-3.7.3.3
52	N+15 Taboo A/N	S	✓	✓						✓		I-3.7.3.3
52	N+15 Taboo A/N	M	✓	✓						✓		I-3.7.3.3
52	N+15 Taboo A/N	W	✓		✓					✓		I-3.7.3.3

ATTC Test #	TEST	D LVL	SUB-TEST							NOTES	TEST PLAN SECTION
			TOA		Stereo CCIR 3	SAP CCIR 3			EO &C		
48	BTSC Audio for N+14 Taboo A/N Stereo/ Mono, & SAP	W			✓	✓			✓	EO&C for 13 TV Sets Stereo/Mono and 9 SAP *This group of tests supplementary to original test plan*	
All above tests use 1080I x 1920 pixel Scan Format											

✓ Data taken in test plan for Grand Alliance system.

‡ = CCIR3 interference levels determined fo all 24 receivers. *Extension from 12+ to 24 receivers supplementary to original test plan*

Transmission Tests

ATTC Test #	TEST	D LVL	SUB-TEST TOV	CCIR 3	POU	POR	POF	RNG	NOTES	TEST PLAN SECTION
259	N-3 Taboo N/A	S	✓*						BER method	I-3.7.3.3
259	N-3 Taboo N/A	M	✓*						BER method	I-3.7.3.3
259	N-3 Taboo N/A	W	✓*						BER method	I-3.7.3.3
260	N-3 Taboo A/A	S	✓*						BER method	I-3.7.3.3
260	N-3 Taboo A/A	M	✓*						BER method	I-3.7.3.3
260	N-3 Taboo A/A	W	✓*						BER method	I-3.7.3.3
29	N-2 Taboo N/A	S	✓*						BER method	I-3.7.3.3
29	N-2 Taboo N/A	M	✓*						BER method	I-3.7.3.3
29	N-2 Taboo N/A	W	✓*						BER method	I-3.7.3.3
30	N-2 Taboo A/A	S	✓*						BER method	I-3.7.3.3
30	N-2 Taboo A/A	M	✓*						BER method	I-3.7.3.3
30	N-2 Taboo A/A	W	✓*						BER method	I-3.7.3.3
33	N+2 Taboo N/A	S	✓*						BER method	I-3.7.3.3
33	N+2 Taboo N/A	M	✓*						BER method	I-3.7.3.3
33	N+2 Taboo N/A	W	✓*						BER method	I-3.7.3.3
34	N+2 Taboo A/A	S	✓*						BER method	I-3.7.3.3
34	N+2 Taboo A/A	M	✓*						BER method	I-3.7.3.3
34	N+2 Taboo A/A	W	✓*						BER method	I-3.7.3.3
261	N+3 Taboo N/A	S	✓*						BER method	I-3.7.3.3
261	N+3 Taboo N/A	M	✓*						BER method	I-3.7.3.3
261	N+3 Taboo N/A	W	✓*						BER method	I-3.7.3.3
262	N+3 Taboo A/A	S	✓*						BER method	I-3.7.3.3
262	N+3 Taboo A/A	M	✓*						BER method	I-3.7.3.3
262	N+3 Taboo A/A	W	✓*						BER method	I-3.7.3.3
All above tests use 1080I x 1920 pixel Scan Format										

✓ Data taken in test plan for Grand Alliance system.

* When only TOV was determined for interference/impairment into ATV, acquisition at TOV was confirmed.

Transmission Tests

ATTC Test #	TEST Effect of Multipath	D I VL	SUB-TEST TOV	POU	POR	POF	RNG	EO&C	NOTES	TEST PLAN SECTION
272, 273, 274, 275, 276, 277, 278 281	Random Noise in presence of Ensembles of 5 Multipaths - 7 Ensembles A - G - No Ensembles for calibration	S	✓*						BER method	I-3.4
266, 267, 268, 269, 270, 271 280	Co-Channel NTSC in presence of Ensembles of 5 Multipaths - 6 Ensembles A - F - No Ensembles for calibration	W	✓*						BER method	I-3.5
285, 286, 287, 288, 289	Strongest Static Echo Rejection - 1 30μs echo w/Ensemble C - 1 5.7μs echo w/Ensemble A - 3 single echoes 15μs 5.7μs & 1μs	S	✓* ✓* ✓*						No noise or Co-Channel added. BER method	I-3.6.2.1
290, 291, 292, 293 294, 295	Strongest Dynamic Echo Rejection - Ensemble A w/1.8μs echo 4 frequencies of phase rotation 0Hz, 0.05Hz, 0.5Hz 5Hz - Single echo 1μs 2 frequencies of phase rotation 2Hz, 5Hz	S	✓* ✓*						No noise or Co-Channel added. BER method plus viewing for 4 frequencies w/1.8μs echo BER method only for 2 frequencies w/1μs echo	I-3.6.2.2
All above tests use 1080I x 1920 pixel Scan Format										

✓ Data taken in test plan for Grand Alliance system.

* When only TOV was determined for interference/impairment into ATV, acquisition at TOV was confirmed.

Discrete Frequency

ATTC Test #	TEST	D	SUB-TEST						NOTES	TEST PLAN
	Discrete Frequency	LVL	TOV	POU	POR	POF	CCIR 3	RNG		SECTION
102 F1 103 F2 104 F3 105 F4 106 F5 107 F6 108 F7 109 F8 110 F9 111 10 112 11 113 12 114 13 115 14 116 15 117 16 118 17 119 18 120 19 121 20 122 21 123 22 232 23 233 24 234 25	25 Discrete Frequencies	W	✓*						BER method	I-3.7.3.4
All above tests use 1080I x 1920 pixel Scan Format										

Objective Tests

ATTC Test #	TEST	SUB-TEST			NOTES	TEST PLAN
	Objective Tests	OBJ	REC	EO&C		SECTION
	Objective Video -Y/C Static Resolution		✓	✓	Archival Recording	I-5
	Objective Video - Dynamic Zone Plate Artifacts		✓	✓	Archival Recording	I-6
	Dynamic Resolution - PIXAR Generated Radial Resolution Chart		✓	✓	Archival Recording	I-5
	H,V,D Transient Response	✓	✓		Archival Recording	I-7
	Temporal Transient Response	✓	✓		Archival Recording	I-7
	Audio/Video/Captioning Latency	✓				I-10
All above tests use 1080I x 1920 pixel Scan Format and 720P Scan Format.						

✓ Data taken in test plan for Grand Alliance system.

* When only TOV was determined for interference/impairment into ATV, acquisition at TOV was confirmed.

Quality Tests

ATTC Test #	TEST Quality Tests	SUB-TEST OBJ	 REC	 EO&C	NOTES	TEST PLAN SECTION
	Video Quality - Basic Input Format 1080I, Display 1080I Input Format 720P, Display 720P (Includes Orig. & New material) Recording includes material required for Digital Specific testing		✓		For GA system testing should be expanded to include additional material created in each claimed input format.	III-7
	Video Quality - Receiver Transconversion Input Format 1080I, Display 720P Input Format 720P, Display 1080I (Includes Orig. & New material) Recording includes material required for Digital Specific testing		✓		For GA system testing should be expanded to include additional material created in each claimed input format. Test images to be evaluated at ATEL were selected by Digital Specific Task Force	III-7
	Concatenation Video • 2 passes thru GA system		✓	✓		I-8
	• 2 passes thru GA system with added key after first pass		✓	✓	1080I only	I-8
	Audio • 2 passes of subjective quality material through GA system		✓	✓	1080I only	I-11
	• 2 passes of subjective quality material through GA system with added voice-over after first pass • 6db reduction • 10db reduction		✓	✓	1080I only	I-11
All above tests use 1080I x 1920 pixel Scan Format and 720P Scan Format except where noted.						

✓ Data taken in test plan for Grand Alliance system.

Digital Specific Tests

ATTC Test #	TEST Digital Specific Tests	D LVL	SUB-TEST TOV	POU	POR	POF	RNG	REC	EO &C	NOTES	TEST PLAN SECTION
	Free-Form Viewing incl new source material and special interoperability material •Input Format 1080I, Display 1080I •Input Format 720P, Display 720P •Input Format 1080I, Display 720P •Input Format 720P, Display 1080I							✓ ✓ ✓ ✓	✓ ✓ ✓ ✓		II-2.1
	Scene Cuts incl stressful images and other new material •Input Format 1080I, Display 1080I •Input Format 720P, Display 720P •Input Format 1080I, Display 720P •Input Format 720P, Display 1080I							✓ ✓ ✓ ✓	✓ ✓ ✓ ✓		II-2.2
	Live 1080I CCD Camera							✓	✓	1080I only	II-2.12
	1035I to 1080I Transconverter							✓	✓	1080I only	II-2.13
58	Threshold Characteristics •Random Noise, RN-I/A - Video	S	✓	✓	✓	✓			✓	BER method vs. Visual Observation (For BER TOV and ACQ are done) 1080I only	I-3.3 and II-2.3
241	- Audio	S	✓	✓	✓	✓			✓	5.01 channels 1080I only	II-2.3
127	•Impulse Noise Universal AC Motor - Video	M	✓*						✓	BER method vs. Visual Observation 1080I only	I-3.3 and II-2.3
All above tests use 1080I x 1920 pixel Scan Format and 720P Scan Format except where noted.											

✓ Data to be taken in test plan for Grand Alliance system.

* When only TOV was determined for interference/impairment into ATV, acquisition at TOV shall be confirmed.

Digital Specific Tests

ATTC Test #	TEST Digital Specific Tests	D I VL	SUB-TEST TOV	POU	POR	POF	RNG	REC	EO &C	NOTES	TEST PLAN SECTION
240	Susceptibility to Random Noise in Video Source Unimpaired Channel - Note Artifacts - Note Noise Enhancement Impaired Channel - Test at 6 impairment levels - Note Artifacts	S S							✓ ✓ ✓		II-2.4
	Motion Compensation Overload								✓		II-2.5
236	Multiple Impairment -Noise & Co-Ch N/A	S	✓*						✓	1080I only	II-2.6
256	Time Varying Channel Impairments	S							✓		II-2.7
	Video Coder Overload							✓	✓		II-2.8
	24fps Film Mode - 60-24 & 24-60fps Transitions								✓		II-2.9
	Video Quality/ Auxiliary Data Tradeoff - Constant (forced) rate method,256kB/s & 1, 2, 3, 4Mb/s 1080I Scan Format - Opportunistic Method							✓ ✓	✓ ✓	Includes reallocation of data packets for Impulse Pay Authorization 1080I only 1080I only	II-2.10 II-2.11
	Long Form Entertainment Program								✓	Video with audio viewing/listening Test at 1dB above TOV interference level 1080I only	III-8
All above tests use 1080I x 1920 pixel Scan Format and 720P Scan Format except where noted.											

✓ Data to be taken in test plan for Grand Alliance system.

* When only TOV was determined for interference/impairment into ATV, acquisition at TOV shall be confirmed.

Interoperability & Packetization Tests

ATTC Test #	TEST Interoperability & Packetization Tests	SUB-TEST OBJ	SUB-TEST REC	SUB-TEST EO&C	NOTES	TEST PLAN SECTION
	Packetization/ Interoperability Tests: - Header/Descriptor Robustness		✓	✓	1080I only	I-9.2
	- Switching between compressed data streams				No repeats for second format	
	Test 1. Switch between same format data streams - follow rules 1080I to 1080I.		✓	✓		I-9.1
	Test 2. Switch between different formats - follow rules 720P to 1080I to 1080P 24fps film.		✓	✓		I-9.1
	Tests 3&4. Ancillary Data Packet replacement test.		✓	✓		I-9.1
	- Compression Layer Interoperability Tests		✓	✓		I-9.3
	- Video (MPEG Compliance)		✓	✓		I-9.3
	- Audio (AC-3 Syntax Compliance)		✓	✓	5 channel & 2 channel	
	- Transport Layer Interoperability Tests		✓	✓		I-9.4
	- MPEG Compliance					
	- Interoperability with ATM Networks		✓	✓	To be tested during the Field Testing 1080I format only	I-9.4
	- Multiple Ancillary Data Services		✓	✓	1080I format only	I-9.4

All above tests use 1080I x 1920 pixel Scan Format and 720P Scan Format except where noted.

✓ Data taken in test plan for Grand Alliance system.

ATV Subjective Audio & Long Form Entertainment Tests

ATTC Test #	TEST ATV Subjective Audio & Long Form Entertainment	SUB-TEST OBJ	REC	EO&C	NOTES	TEST PLAN SECTION
	ATV Multichannel Audio (Subjective Audio Tests)				Test Audio Transparency	
	• 5.1 Channel		✓	✓		V
	• 2 Channel Mix-down		✓	✓		V
	• Stereo In/Out		✓	✓		V
	Long Form Entertainment Program		✓	✓	Video with audio viewing/listening test 1080I only	III-8
All above tests use 1080I x 1920 pixel Scan Format and 720P Scan Format except where noted.						

✓ Data taken in test plan for Grand Alliance system.

Legend:

TOV	-	Threshold of Visibility	RN	-	Random Noise
POU	-	Point of Unusability	I/A	-	Impairments into ATV
POR	-	Point of Reception	I/N	-	Impairments into NTSC
POF	-	Point of Failure	CCIR 3	-	CCIR 3 Reference Level
RNG	-	Ranging	CCIR 4	-	CCIR 4 Reference Level
EO&C	-	Expert Observation and Comment	REC	-	Recording
W	-	Weak signal level	OBJ	-	Objective
M	-	Moderate signal level	N/N	-	NTSC into NTSC
S	-	Strong signal level	A/N	-	ATV into NTSC
UP	-	Upper Adjacent Channel	N/A	-	NTSC into ATV
LO	-	Lower Adjacent Channel	A/A	-	ATV into ATV
CO	-	Co-Channel	P	-	Progressive
BER	-	Bit Error Rate	I	-	Interlace
H,V,D	-	Horizontal, Vertical, Diagonal	D LVL	-	Desired signal level
N+or-n	-	A Taboo channel where N is the desired channel and N +or-n is the Taboo channel.			
TOA	-	Threshold of Audibility			

Part VII

Record of Test Results

for

digital HDTV
Grand Alliance System

from

Field Tests

Terrestrial Broadcast Field Tests

Conducted by

Association for Maximum Service Television
Public Broadcasting Service
(July 25 - August 23, 1995)

Cable Field Tests

Conducted by

Cable Television Laboratories, Inc.
(July 28 - August 29, 1995)

Terrestrial Broadcast Field Tests

Conducted by

Association for Maximum Service Television
Public Broadcasting Service
(July 25 - August 23, 1995)

Association for Maximum Service Television
1776 Massachusetts Avenue, N.W., Suite 310, Washington, D.C. 20036 • 202/861-0344

Public Broadcasting Service

TABLE OF CONTENTS

TESTING TERRESTRIAL BROADCAST TRANSMISSION OF THE *DIGITAL* HDTV GRAND ALLIANCE SYSTEM PROTOTYPE

Executive Summary

Following completion of laboratory testing at the Advanced Television Test Center (ATTC), the Grand Alliance[1] (GA) made available for terrestrial broadcast and cable field testing two laboratory-built prototype HDTV[2] systems. One system was tested at ATTC. The second system was certified by GA to be the equivalent of the system tested. This report describes the results of the terrestrial broadcasting tests.

A September 16, 1994 report (SS/WP2-1354) described the results of field testing the GA HDTV transmission subsystem. In the absence of a complete system permitting the observation of pictures and sound, the 1994 tests used as a criterion for satisfactory HDTV reception a bit-error rate (BER) of 3×10^{-6} errors per second. That BER, corresponding also to a segment error rate (SER) of 2.5 packet errors per second, had been determined in the laboratory to produce the threshold of visibility (TOV) of video impairments. In the 1995 tests reported herein, the availability of a complete system permitted judgments to be made on subjective observation of pictures and sound, in addition to objective measurements of such parameters as signal strength, signal-to-noise ratio, equalizer performance and SER.

Locations selected for the 1995 tests were a subset of locations used for test measurements in 1994. Unlike the 1994 selection of 199 sites designed to provide a proper statistical sample of performance throughout the test area, the subset of 40 sites was selected deliberately to include a large number of locations where the 1994 testing on channel 53 suggested marginal, or even submarginal HDTV (and NTSC) performance. Since channels 6 and 53 performances did not track at all times, the subset of test locations can be considered to be somewhat arbitrary insofar as channel 6 is concerned. Furthermore, since channel 6 could be used at only 169 of the 199 channel 53 sites because of interference to cable reception on channel 6, some locations used in 1995 did not have 1994 data for companion channel 6. The 40 sites were supplemented with 10 sites at residences where tests were performed both within the residence, using a set-top antenna, and outdoors, adjacent to the residence, using the usual mast-mounted antenna.

The 1995 terrestrial broadcast field testing reported herein began on July 25 and was completed on August 23. The tests were conducted using the same facilities near Charlotte, North Carolina, as employed in 1994. Again, the NTSC transmitted peak visual effective radiated powers (ERP) on channels 6 and 53 were one-tenth of the maximum allowed by FCC rules, and the average ATV ERP was approximately one-sixteenth of (12 dB below) the NTSC peak visual ERP. The tests were conducted by the Advanced Television Field Test Project staff in Charlotte, supported by representatives from the Grand Alliance, who provided the services to assure continued proper operation of the system prototype. During a majority of the tests, a representative of the FCC served as an observer.

---

[1] A consortium including AT&T, David Sarnoff Research Center, General Instrument Corporation, Massachusetts Institute of Technology, Philips Electronics North America Corporation, Thomson Consumer Electronics and Zenith Electronics Corporation.

[2] In this report, HDTV and ATV are used interchangeably.

Results of the 1995 field testing of the GA complete system prototype support, for UHF channel 53, a conclusion that the more complete testing at 199 sites using BER as the criterion for satisfactory HDTV performance provided an acceptable measure of the reliability of the GA system for digital terrestrial broadcasting. Logic certainly suggests that a similar conclusion is applicable to channel 6 performance; however, because of the small sample size and the channel 6 interference problems in Charlotte that plagued both the 1994 and 1995 tests, no conclusion can be reached based solely on test results. Additional field testing would be most desirable for establishing, without question, the performance reliability of digital television terrestrial broadcasting on VHF channels.

The full system prototype tests showed, as indicated by the 1994 transmission subsystem testing, that satisfactory digital HDTV reception is available more widely than satisfactory analog NTSC reception. Even where objective measurements of SER indicate the probability of momentary impairment of the signal, subjective observation of picture and sound fail to detect impairment. The 1995 subjective assessments of channel 53 performance correlate very well with the 1994 data based on BER. That correlation does not hold for channel 6, but as noted above, sample size and interference effects prevent a proper channel 6 analysis.

An objective measurement that should permit reliable prediction of satisfactory HDTV service is field strength. As shown in the report, subjective assessment of video and audio correlated very well with field strength. As expected, with signal strength at or below that which laboratory testing had indicated to be the limit of HDTV service, only two of the seven sites where such signal strength was encountered demonstrated subjectively satisfactory channel 53 service. On the other hand, only one site (out of a total of 15), where the signal strength was weak but above the threshold, did not have subjectively satisfactory HDTV service. The 28 remaining sites[1] with moderate or strong signal strength, all had subjectively satisfactory HDTV service, thus indicating that field strength can be used as a reliable predictor for satisfactory HDTV service.

In brief, the 1995 field testing of the GA full system prototype supports the conclusion of the 1994 transmission subsystem testing that HDTV service will be available where NTSC service is presently available, and in many instances where NTSC service is unacceptable.

---

[1] The total number of sites amounts to 50 because the 10 residential sites were added to the 40 sites selected as described.

I. INTRODUCTION

A September 16, 1994, report (SS/WP2-1354) described the methodology employed and results obtained from field testing of the transmission subsystem portion of the HDTV system developed by the Grand Alliance, a consortium including AT&T, David Sarnoff Research Center, General Instrument Corporation, Massachusetts Institute of Technology, Philips Electronics North America Corporation, Thomson Consumer Electronics and Zenith Electronics Corporation. The absence of a complete system precluded the delivery of pictures and sound. In place of the normal video and audio for television transmission, a signal suitable for bit error rate (BER) measurements was transmitted. The criterion for establishing whether acceptable HDTV reception could be expected at a site was a BER of 3×10^{-6} or better. Laboratory tests had determined that level to constitute the threshold for satisfactory service.

In the 1994 tests, measurements were made at 199 locations representing a wide range of propagation conditions. Although the Federal Communications Commission had granted authority to deliver test transmissions on both UHF channel 53 and VHF channel 6, the extent of measurements using channel 6 was curtailed because of complaints of interference to cable viewers. However, channel 6 measurements were conducted at 169 of the 199 sites. Channel 53 was measured at all 199 locations.

Upon completion of laboratory testing of the Grand Alliance prototype ATV system at the Advanced Television Test Center (ATTC) in June, 1995, the tested prototype and a second unit certified by the Grand Alliance to be the equivalent of the tested system were made available for field testing. One complete system prototype unit was used for the testing of system performance when subjected to the terrestrial transmission environment. The second unit was dedicated to the determination of system performance when delivered by cable.

Actual testing of terrestrial transmission performance started on July 25, 1995, after an initial period for system installation and check out, and calibration. The Phase I testing, reported herein, was completed on August 23. The purposes of the Phase I test program were two-fold: (1) to provide a check of the validity of laboratory test results when the system is operated under real world propagation conditions, and (2) to determine whether the results of the 1994 tests, using bit error rate as the criterion for performance, were a reliable measure of the acceptability of the Grand Alliance system for terrestrial broadcasting.

In the course of the Phase I testing, observations and measurements were made on channels 53 and channel 6 transmissions at a total of 40 radial, cluster and grid locations. Channel 6 observations and measurements were made at eight of the ten in-home locations, where observations and measurements were also made on channel 53 transmissions. At the last two in-home locations, measurements on channel 6 were discontinued because of reports of interference in local cable homes, primarily from the NTSC signal. Despite the fact that channel 6 data are available for 56 locations (at residences, observations and measurements are made both in-home and outdoors, adjacent to the residences), a complete analysis of channel 6 results is not possible. At the 56 sites, only 12 were free of interfering sources. At 31 locations, impulse noise was encountered from either or both power lines and atmospheric sources. At 13 locations, serious interference was encountered from noncommercial educational FM stations, and co-channel interference was encountered at 17 locations. (The foregoing figures total to more than 56 because interference from multiple sources was encountered at some locations.)

FM interference encountered in Phase I testing was worse than that experienced in previous testing performed in 1994. Serious FM interference had been experienced previously,